COMPARATIVE ANALYSIS OF CRIME TYPE PREDICTION

R. Ananda DhanaLakshmi[1], Dr. Grasha Jacob[2]
[1]Research scholar, Register Number:18121172162002
[1]Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli 627 012, Tamil Nadu, India
[2]Associate Professer & Head
[1,2] Dept of Computer Science
Rani Anna Govt College for women ,Tirunelveli 627 008,Tamil nadu, India
grasha.ananthi@gmail.com

ABSTRACT: Crime analysis and prediction is a logical approach for analysing and identifying different patterns, relationships and trends in crime. The system can predict the area which have high probability for crime occurrence and indicate crime prone areas. It will be useful for the law enforcement agencies to speed up the process of solving crimes with the increasing computerized systems and with the help of crime data analysts. Here we have an approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Analysis of police data createawareness that lets officers track criminal activities, predict the incidents, effectively deploy resources and solve cases faster. In this paper crime analysis is done by k-means,Spectral clustering and Agglomerative clustering on the crime dataset.

Key Words: Crime-patterns, data clustering, data mining, k-means, fuzzy c, law-enforcement.

1. Introduction

Crimes are a common social badly behaved affecting the quality of life and the economic growth of a society. It is measuredas an essential factor that determines whether or not people move to a new city and what places should be avoided when they travel. With the increase of crimes, law enforcement agencies are continuing to demand advanced geographic information systems and new data mining approaches to improve crime analytics and better protect their communities.Although crimes could take place everywhere, it is common that criminals work on crime opportunities they face in most familiar areas for them. By providing our approach to identify the most criminal hotspots and find the type, location and time of committed crimes, we hope to increase people's awareness regarding the unsafe locations in certain time periods. Therefore, our solution can help people stay away from the locations at a certain-time of the day along with saving livesIn addition, having this kind of knowledg e would help people to get   more   out of their living place selections. On the other ha nd, police forces can use this answer to increas e the level of crime statement of what will take place in the future and putting a stop to. in addition, this would be useful for police resour ces a thing or amount put to one side.

Crime analyses is an emerging field in law enforcement without standard definitions. This makes it difficult to determine the crime investigates focus for agencies that are new to the field. In some police departments, what is called "crime analysis" consist of mapping crimes for command staff and producing crime statistics. In other agencies, crime analysis might mean focusing on analyses various police reports and suspect information to help investigators in major crime units.

2. Literature Survey

Data mining in the study and analysis of criminology can be categorized into main areas, crime control and crime suppression.

De Bruin et. al. [1] suggested a framework for crime movements using a new distance

measure for comparing all individuals based on their outlines and then clustering them accordingly.

Manish Gupta et. al. [2]. Rendered the existing systems used by Indian police as e-governance initiatives and also offers an interactive query based interface as crime analysis tool to assist police in their activities. He proposed interface which is used to extract useful information from the vast crime database maintained by NCRB and find crime hot spots using crime data mining techniques such as clustering etc.

NazlenaMohamad Ali et al.[3] applied on a development of Visual Interactive Malaysia Crime News Retrieval System (i-JEN) and describe the approach, user studies and planned, the system architecture and future plan. Their main objectives were to make crime-based event; investigate the use of crime based event in improving the classification and clustering; make an interactive crime news retrieval system; visualize crime news in an effective and interactive way; integrate them into a usable and robust system and evaluate the usability and system performance and the study will contribute to the better understanding of the crime data consumption in the Malaysian context as well as the developed system with the visualization features to address crime data and the eventual goal of combating the crimes .

SutapatThiprungsri [4] applied the application of cluster analysis in the accounting domain, particularly inconsistency detection in audit. The purpose of his study is to inspect the use of clustering technology to automate fraud filtering during an audit. He used cluster analysis to help auditors focus their efforts when evaluating group life insurance claims.

A. Malathi et al.[5] suggested the use of missing value and clustering algorithm for a data mining approach to help predict the crimes patterns and fast up the process of solving crime. Malathi. A et. al.[6] introduced a clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyse the city crime data from Police Department. The results of this data mining used to prevent crime for the coming years.

Dr. S. SanthoshBaboo and Malathi. A [7] Developed a crime analysis tool for Indian scenario using different data mining techniques these techniques help to the law enforcement department to efficiently handle crime investigation. The proposed algorithmhelps to analyse crime data to identify actionable patterns and trends.

Kadhim B. Swadi Al-Janabi [8] developed a proposed framework for the crime and criminal data analysis and detection using Decision tree Algorithms for data classification and Simple K Means algorithm for data clustering. The paper tends to help in discovering type and fashions, making predictions, finding relationships and possible explanations, Finding relationship between criminal networks and identifying possible suspects.

AravindanMahendiran et al. [9] Described myriad of tools on crime data sets to mine for information that is hidden from human perception. Through the art visualization techniques it identifies all the patterns through our algorithms. That helps the law enforcement departments to channelize their resources accordingly.

SutapatThiprungsri[10] Suggested the possibility of using clustering technology for auditing. Automating fraud filtering can be of great value to continuous audits. The objective of their study is to examine the use of cluster

analysis as an alternative and innovative anomaly detection technique in the wire transfer system.

3. Proposed Model

Our aim is to predict the Crime type in a particular location. It creates the awareness among people.The data was collected from the city of Chicago data portal. The city of Chicago have the highest population and highest crime rate. From the dada set only 9 attributes are selected. For example crime number, crime date, latitude, longitude. And data was pre-processed.
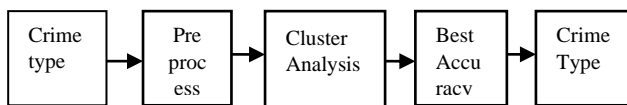

Figure 2: Architecture of proposed method

3.1 K-Means Clustering

K-means algorithm is an iterative algorithm that tries to partition the dataset into '*K'* predefined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

1. Initially, k- the number of clusters

2. The first step is the chose a set of K instances as centres of the clusters.

3. Next, the algorithm reflects each instance and assigns it to the cluster which is closest one.

4. The cluster centroids are recalculated. The whole cycle of re-loop or each instance assignment.

5. This process is iterated.

3.2 Spectral Clustering

Spectral Clustering is a growing clustering algorithm and performed better than many traditional clustering algorithms in many occurrences. It treats each data point as a graph-node and thus transforms the clustering problem into a graph-partitioning problem. Spectral clustering is a family of methods to find K clusters using the eigenvectors of a matrix. Typically, this matrix is derived from a set of pairwise similarities Sij between the points to be clustered. This task is called similarity based clustering, graph clustering, or clustering of crime data.

3.3 Agglomerative Clustering

The agglomerative clustering is the most common type of hierarchical clustering used to group objects in clusters based on their similarity. It is known as *AGNES* (*Agglomerative Nesting*). The algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects. The result is a tree-based representation of the objects, named *dendrogram*.

4. Performance metrics

Clusrers	Sensitivity	Specificity	Precision	Accuracy
K-Means	**0.64**	**0.58**	**0.56**	**0.87**
Spectral	0.38	0.69	0.31	0.75
Agglomerative	0.78	0.52	0.53	0.75

The methodology here is through an open source data set was collected from city of Chicago data portal (www.data.cityofcicago.org) . The data was gathered from Chicago because it has the largest population and high crime rate. It comprises 2, 68,642 of violent crime in Chicago in the year 2018.Data integration process method is meticulously followed.First to avoid different attributes that combined the key attributes, the date, time, and the remaining fields in different granularities were analysed. Then again the army time system for date and time was pursued. Subsequently the hour part without paying attention to the minutes to obtain more of frequent patterns was deemed. As there are some missing values in our data set, removal of missing values was imperative to obtain an appropriate crime data set. Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Often the data is unstructured, inconsistent, which has missing values, and lack in certain behaviour or trends that is prone to many errors. So the data need to be cleaned. Cleaning means removing the missing value and noisy data. First translation was applied to thedata format. The non-crime points were generated by mapping the location of daily crime occurrence with the nearest points.

4.1 Experimental and result

Table 1. Results of performance

4.2 Clustering Metrics

Cluster analysis metrics is finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters.Let 'N' be the number of observations in a sample. Let $\propto$ to be the number of observation pairs with the same labels and located in the same cluster, and let 'b' to be the number of observations with different labels and located in different clusters. The Rand Index can be calculated., this index is a measure of distances between different sample splits. The values in the [-1,1] range. Negative values indicate the independence of splits, and positive values indicate that these splits are consistent. From the metrics k means clustering gives the best Accuracy.
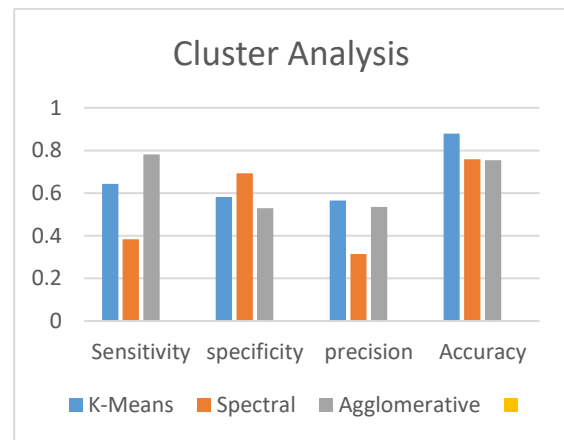

Figure 3: performance chart for the above table1

5. Conclusion

The increase in the rate of crime in today's world is a very important thing that needs to be curbed for a safer environment. In this paper, we have analysed the crime prediction using the data mining techniques.The crime rate can be predicted by providing the appropriate input and also the changing crime patterns. Crime patterns cannot be static since patterns change over time, so different clustering techniques like K-Means; Spectral Clustering etc. are used to identify the changes in crime patterns and provide better accurate results.

REFERENCES

[1] De Bruin ,J.S.,Cocx,T.K,Kosters,W.A.,Laros,J. and Kok,J.N "Data mining approaches to criminal carrer analysis ,"in Proceedings of the Sixth International Conference on Data Mining (ICDM"06) , **(2006)**,Pp. 171-177

[2] Manish Gupta1*, B.Chandra1 and M. P. Gupta1 Crime Data Mining for Indian Police Information System **(2007).**

[3] NazlenaMohamad Ali1, Masnizah Mohd2, Hyowon Lee3, Alan F. Smeaton3, Fabio Crestani4 and ShahrulAzmanMohd Noah2 , "Visual Interactive Malaysia Crime News Retrieval System" **(2010).**

[4] SutapatThirprungsri Rutgers University .USA ,2011 Cluster Analysis of Anomaly Detection in Accounting Data : An Audit Approach 1

[5] A.Malathi ,Dr.S.SanthoshBaboo. D.G. VaishnavCollege,Chennai , "Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters", **(2011).**

[6] Malathi.A1 ,Dr.S.SanthoshBaboo 2 and Anbarasi . A 31 Assistant professor ,Department of Computer Science ,Govt Arts College ,Coimbatore , India . 2 Readers , Department of Computer science , D.G. VaishnavCollge ,Chennai , India , "An intelligent Analysis of a city Crime Data Using Data Mining", **(2011).**

[7]Malathi, A;SanthoshBaboo S,"An Enhanced Algorithm to Predict a Future Crime using DataMining" , **(2011)** .

[8] KadhimB.Swadi al-Janabi . Department of Computer Science. Faculty of Mathematics and Computer Science University of Kufa/Iraq , "A Proposed Framework for Analyzing Crime DataSet using Decision Tree and Simple Kmeans Mining Algorithms",(**2011)**.

[9] AravindanMahendiran, Michael Shuffett, SathappanMuthiah, Rimy Malla, Gaoqiang Zhang, "Forecasting Crime Incidents using Cluster Analysis and Bayesian Belief Networks",(**2011)**.

[10] Sutapat Thiprungsri, "Cluster Analysis for Anomaly Detection in Accounting Data :An Audit Approach1",(**2012)**.